Notes from NTAPAA 3/11/05 Meeting on Test Design
Brian Gong, Center for Assessment

Purposes for Kentucky Core Content Tests (KCCT)

The National Technical Advisory Panel on Assessment and Accountability (NTAPAA) appreciated the clear explication from Commissioner Wilhoit of the main purposes for the Kentucky Core Content Tests (KCCT). They noted that school accountability and student reporting for lower stakes (e.g., inclusion as part of local classroom grading) were consistent with Kentucky's current system. Higher student stakes (e.g., mandatory placement in remedial summer school, promotion or graduation decisions) would require somewhat different designs for the writing assessment discussed the previous day (3/10) or for the discussion of test design for other content areas covered today (3/11).

Test Design for the KCCT in Reading and Math

NTAPAA noted test design would involve technical and policy decisions that strike a balance between content coverage and comparability at the student level. NTAPAA members tended to emphasize the advantages of a simpler test design that would speed processing, allow for greater student comparability, and allow for student subscore reporting. (e.g., a more focused design in that content slated for assessment at each grade level would include only the most important/relevant content/processes for that grade.) Based on a general view of the revised Core Content having approximately 5 strands and 20 bullet statements, panel members felt it likely that reasonable content coverage could be achieved with the revised Core Content over a period of a few years at most, using a single common form each year – a design that would need checking by specific analyses. It was noted that for each of the 20 bullet statements, a number of items could be developed that assess various aspects and levels of understanding of the bullet. We need to recognize that each bullet probably covers a range of content/cognitive processes. (The discussion on Core Content and cognitive complexity the previous day (3/10) also cautioned against assuming that performance on one item would generalize broadly.) NTAPAA members emphasized the advantages for test design of having a very focused content specification.

NTAPAA focused on a test design of a common core, or single form, for all students. A different, comparable (but not strictly parallel) form would be administered each year. Student scores (overall score for the content area plus some subscores) would be based solely on the common core. This common core could be supplemented with matrix items. The matrix items could be used for school accountability, although this might require more complex processing, scaling, and accountability usage. They felt that use of matrix items for student scores was also feasible, given a substantial common core, but there would be drawbacks in more complex processing and provision of student subscores. The issue was raised of using matrix scores for school accountability—as has

been done previously in Kentucky's school accountability program—but that issue was not discussed in depth.

NTAPAA members generally saw few advantages in shorter constructed response formats, based on the formats with which they were acquainted. NTAPAA noted that Kentucky's test design should retain substantial Open-Response (OR) items of the length, complexity, and format currently used if the state continues to value those types of performances. When pressed about possibly shortening the testing time, an NTAPAA member gave the opinion that the total number of OR items could be as few as three per form, but noted that was really a policy decision, not a technical decision; the other NTAPAA members did not weigh in on this issue of a specific number of OR items, although another NTAPAA member noted that in his experience, OR items in writing have contributed to improved instruction and learning in writing.

NTAPAA recommended developing and pre-equating several forms, if possible. Development should include enough to release a substantial number of items.

NTAPAA member Bob Linn noted that this design could be all custom for every grade level, or an all augmented norm-referenced test (NRT) for every grade level. Linn felt that an augmented NRT could provide several advantages, including provision of a vertical scale and perhaps lesser cost. Other NTAPAA members said that they felt alignment was very important, and expressed some skepticism that an NRT would meet the state's content standards adequately. All agreed that the state should be sure that any proposal based on an NRT included more than the vendor's assertion that the test would be aligned with Kentucky's content standards. In addition, issues such as equating and differential sensitivity to improvement in performance across the NRT/custom portions would need to be addressed. Bob Linn raised the possibility of starting with an NRT core and transitioning to a fully custom test, and pointed out several disadvantages to this approach as well. John Poggio mentioned the possibility of administering a minimal survey form in limited grades to get student-level NRT information.

Test Design Across Grade Levels

NTAPAA members agreed that there should be a single test design across all grade levels for reading and math. There should not be a mix of augmented NRT and KCCT of very different designs, as will be the case for 2006 (and probably 2007).

Test Design for KCCT in Science and Social Studies

NTAPAA members did not extensively discuss the issues of test design for science and social studies. They did note that the test design for reading and math could be extended to science and social studies if the decision were made to assess a single grade of content in each grade span. Presumably more matrix items and/or forms would be needed if KDE continues to assess a grade span of content each once in elementary, middle, and high school, e.g., assess the content for grades 3-5 for a test administered in grade 5, and content for grades 6-8 for a test administered in grade 8.

Test Design for Arts & Humanities and Practical Living/Vocational Studies

NTAPAA recognized and discussed the difficulties in assessing arts and humanities and practical living/vocational studies. These difficulties included broad content, lack of agreement on focus (e.g., broad appreciation versus development of individualized performance skills in perhaps one area of the arts), and inherent limitations of a large-scale assessment for some of the constructs (e.g., fitness). Although NTAPAA discussed several possible approaches to assessments for these content areas, they acknowledged they were not able to recommend satisfactory alternatives in terms of test designs within the limited time allocated for discussion. They were sympathetic to considering assessment not accomplished through on-demand, large-scale work—such as the proposal to review individual student fitness plans for assessing practical living mentioned by Commissioner Wilhoit—but acknowledged the operational difficulties and the large degree of support needed for such a change.

Scaling for Monitoring Student Longitudinal Growth

NTAPAA discussed various options for scaling for monitoring student longitudinal growth for the reading and math assessments. They pointed out that care would need to be taken with vertically moderated standards or with a vertical scale (single scale across the entire grade span or single scales across fewer grades). In the limited time available, detailed issues of scaling were not raised, such as the test design implications of a vertical scale or the interpretation of proficiency for a grade-level test that included substantial "off-grade" items to link to adjacent grades. NTAPAA did not recommend a particular scaling method, and agreed that the Request for Proposals (RFP) should seek demonstration of competence from the vendor because a specific longitudinal scaling approach could only be specified after many other details of test design had been specified.

HumRRO staff (Gene Hoffman and Art Thacker) shared some of HumRRO's experience in identifying and addressing issues of creating a vertical scale from state data. HumRRO also pointed out that scales and performance standards would likely change rapidly for some grades as instruction moved from a concentration in a few grades to all grades in reading and math.

Formative/Diagnostic Assessment

NTAPAA was very supportive of KDE's intent to have formative/diagnostic and other local assessment that would complement the formal student assessments used for school accountability. They commented that it would be important to have considerable coordination between the two systems; they mentioned that in their view, a single vendor might be better able to produce such tightly coordinated materials, although the actual delivery might be by another vendor.

Future Interaction Around the RFP

Commissioner Wilhoit said that it was KDE's intention to bring the RFP in more detail to NTAPAA in June prior to the KBE meeting. He hoped the RFP would be approved by KBE for public release in July or August 2005. Commissioner Wilhoit also raised the possibility of asking NTAPAA to help evaluate the proposals.